The Design Effects of Voting Advice Applications: Comparing Methods of Calculating Results

Tom Louwerse
University of Leiden
tlouwerse@fsw.leidenuniv.nl

Martin Rosema
University of Twente
m.rosema@utwente.nl

Abstract

In election times more and more voters have consulted Voting Advice Applications (VAAs), such as smartvote in Switzerland, Wahl-O-Mat in Germany or EU Profiler across EU Member States. The potential impact of these online tests on election outcomes is substantial and hence it is important to study the effects of their design. This paper focuses on an important element of the design, namely the method used to calculate the match between voters and parties. More specifically, we examine the use of alternative (implicit or explicit) spatial models and metrics. The analyses are based on the actual answers given by users of one of the most popular VAAs in Europa, StemWijzer in the Netherlands. The results indicate that the advice depends strongly on the spatial model adopted. A majority of the users of StemWijzer would have received another advice, if another spatial model had been used. These findings have important implications for the design of future VAAs.

Section: Internet and Politics: Bridging Current Research and Outlining Future Directions.
Panel: Voting Advice Applications under Scrutiny – Assessing their Impact on Elections.
In election times many voters turn to tests on the internet that tell them which party or candidate provides the best match. These tests often have fancy names, such as *smartvote* in Switzerland, *Wahl-O-Mat* in Germany, and *StemWijzer* in the Netherlands. Especially in political systems where voters are faced with a multitude of parties or candidates they have become quite popular. In countries like Finland, Switzerland, and the Netherlands more than one third of the electorate consulted a Voting Advice Application (VAA) before casting their ballot in the national elections. Typically, these tests ask individuals to indicate their personal opinions about a series of policy items, which are then compared with the positions taken by the competing parties or candidates. The outcome is a screen that presents the match or mismatch between the individual and the alternative parties or candidates, which may be conceived of as an advice to vote for the party with the best match.

The method that VAAs have adopted to determine the degree of match between voter and party differs. One important element of the method is the type of spatial model that is used to calculate the match and to present the advice. For example, *Kieskompas* in the Netherlands plots the voter and parties in a two-dimensional political space. The implicit message is that the party closest in space provides the best match. The Swiss *smartvote* analyses policy preferences in terms of eight distinct dimensions and presents the results with a special diagram, which looks somewhat like a spider web. Even VAAs that do not present the results using a spatial framework, but for instance use bar charts for each party, often adopt a particular spatial model for making the underlying calculations. We expect that the method chosen, in particular the type of spatial model used, may affect the party that appears to be the best choice. In this paper we therefore examine whether, and to what extent, the choice for a particular spatial framework influences the outcome of the tests. Furthermore, we also explore a related issue, namely the effect of the metric used to calculate distances.
The spatial framework adopted by VAAs to determine the match is not the only part of the method that may impact the results of the tests. Another key element is the selection of statements. It may well be that if an existing test would have used a different set of policy items, users would have been given another advice. Indeed, Walgrave, Nuytemans and Pepermans (2009) found that statement selection has a profound impact on the match observed between party and voter. They ran a large number of simulations in which a subset was taken from a larger pool of items on which the positions of parties and users were known. The findings indicated that the selection of statements not only affected the outcome at the individual level for voters, but also at the aggregate level for parties: with certain sets of items particular parties turned out to provide the best match with many more voters than with other sets of items. To date, this is the only published study (to the best of our knowledge) in which the impact of the method employed by VAAs on the outcome of the test has been systematically examined. The research on VAAs available, which is still not that much, has focused primarily on the type of individuals that visit these websites (Wall et al. 2009; Fivaz and Nadig 2010), whether the advice matches pre-existing preferences (Wall et al., 2009), the type of parties that perform well in VAAs (Kleinnijenhuis et al. 2007; Ramonaitė 2010), the effects on electoral turnout (Fivaz and Nadig 2010), and the effects on party or candidate choice (Kleinnijenhuis et al. 2007; Walgrave, Van Aelst and Nuytemans 2008). The impact of the design of VAAs on their outcomes has not (yet) been at the heart of the research agenda.

There are several other elements of the design of VAAs than statement selection that merit attention from researchers. One of them is the method used to determine the positions of parties with respect to the items included in the test. Some other examples are the number of answer categories used to respond to statements, the way “don’t knows” and “neutral positions” are handled, and the weighting procedures applied – often users (and sometimes
parties) can indicate which topics they consider more or less important, which is then taken into account. The research presented in this paper fits this broader field of study that we expect will develop in the coming years. The aim of the paper is to contribute by examining the relevance of the spatial framework that VAAs adopt and the precise procedure used to determine the overall match between user and party, which is the basis for the advice provided.

It is highly relevant to understand the effects of the methods employed by VAAs, because these tests may have a substantial impact on the outcome of elections. The last two decades in many western democracies VAAs have been introduced, so in more and more countries voters have the availability to visit the internet and receive advice for an upcoming election (Cedroni and Garzia 2010). Moreover, many voters take advantage of this opportunity and complete such a test in the weeks or days, or sometimes even hours, before going to the polling station. The advice given to voters appears to have a significant impact on their vote choice, especially among floating voters (Ladner, Felder and Fivaz 2010; Walgrave et al. 2008; Ruusuvirta and Rosema 2009). This means that the potential impact of VAAs on the outcome of the election is substantial and hence it is important to study the effects of their design (Wagner and Ruusuvirta 2009; Walgrave et al. 2009).

We examine to what extent the spatial framework adopted affects the advice given to voters. On the basis of an overview of methods employed by VAAs, we distinguish between four alternative spatial models. We determine how often the use of a different spatial framework would lead to another advice. The basis for our analysis is the actual answers given by millions of users of one of the most popular VAAs in Europe, StemWijzer in the Netherlands. In 2010 this test was completed almost 4.2 million times, which roughly corresponds with 40 percent of the eligible voters. The results to be presented below indicate that the spatial model matters a great deal. A majority of users would have been given
another advice, had another spatial model been adopted. The metric used to calculate
distances also had an impact, but not that strongly. These findings have important
implications for the interpretation of the results by voters and for the design of future VAAs.

The paper is structured as follows. In the next section we discuss the methods
employed by different VAAs to calculate the match between voters and parties and to present
the advice. We pay special attention to StemWijzer in the Netherlands, because we use data
from this test in the empirical analysis. Next we outline four alternative spatial models that
are compared in this study and we discuss the corresponding methods to calculate the match
between voters and parties. Reflecting the standard procedure used in VAAs, our analysis
then proceeds in two steps. First, for each voter the degree of match with each party is
calculated. Second, for each voter the party with the best match is determined. In the results
section we demonstrate that the choice for a particular spatial framework affects the outcome
at two levels: for individual voters there are differences in terms of which party appear to
provide the best match, and for political parties there are differences in terms of how often
they appear to provide the best match with voters. In the final section we discuss how
developers of voting advice applications can take our findings into account.

**Voting Advice Applications and their methods**

One of the earliest Voting Advice Applications (VAAs) is StemWijzer, which was developed
for educational purposes by the Dutch non-profit organisation Instituut voor Publiek en
Politiek.\(^1\) The first version of StemWijzer was created in 1989 as a paper and pencil test
meant for high school education (De Graaf 2010).\(^2\) The test was also available on a floppy
disk, but only fifty copies were sold in 1989 and about one thousand in 1994. The paper
version was more popular, but it did not come close to the usage figures that would later be
reached. In 1998 the test became available on the internet and 6,500 voters completed it
online. In subsequent national elections the number of users grew rapidly: in 2002 the website was consulted 2.0 million times and in 2006 this number increased to 4.8 million. This amounts to about one third of the total population and almost forty percent of those eligible to vote.\(^3\) In 2010 the figure dropped to 4.2 million, probably because several other organisations developed their own online test. The most popular alternative is *Kieskompas*, which in 2006 was consulted about 1.5 million times (Kleinnijenhuis et al., 2007: 42; figures for 2010 are not known). *StemWijzer* has also developed tests for elections at the municipal and provincial level, as well as for elections for the European Parliament. There are several spin-offs in other countries, like *Wahl-O-Mat* in Germany and *Politarena* in Switzerland, which originate in cooperation between the Dutch developers and non-profit organisations in those countries (Marschall and Schmidt 2010; Ladner et al. 2010).

The design of *StemWijzer* has undergone some changes throughout the years. The first version contained 60 ideological statements derived from election manifestos (De Graaf 2010). In later tests the number of statements was reduced to 30, which were taken from a long list of about 50 statements for which party positions would be determined. The basis for party positioning is election manifestos and consultations with the headquarters of political parties. When compiling the final list of items, important criteria are that the statements involve some controversy (there should be parties in favour as well as against and most parties should take a position) and that the list as a whole provides sufficient opportunity to distinguish between any pair of parties. For each statement parties as well as users have three alternative positions: agree, neutral (or ‘neither’), and disagree. Additionally, users have the possibility to indicate no opinion and skip the item.

In the first paper and pencil test the statements were used to plot users and parties on a left/right continuum. In later years, however, the use of this simple spatial model was abandoned. In all online versions of *StemWijzer* the match between party and user has been
determined without the explicit use of a spatial model. Between 1998 and 2006, for each party the mismatch was calculated by awarding 2 points if party and user take opposite stands on a statement (agree versus disagree) and 1 point if either of them is neutral whereas the other agrees or disagrees (De Graaf 2010). Furthermore, users could identify any number of statements that they consider more important and these items are then given extra weight (points are doubled). The match with each party is calculated as a percentage of agreement (percentage of maximum points distracted from 100 percent). On the final screen the user is told which party provides the best match. Additionally, parties are ranked according to the degree of match and bar charts are used to visualize this (see Figure 1a).

A key characteristic of the method employed by StemWijzer is that it does not explicitly use a spatial framework for making calculations or presenting the advice (except for the original paper and pencil test). However, the method may still be conceived of in those terms. It basically corresponds with a spatial model in which each item represents a separate dimension. The metric used is city block distance, which deviates from the Euclidean distance measure that is more common in spatial models. Since 2006, StemWijzer has used an even simpler method: one point is awarded to a party for each match between the opinion of the user and the position of the party, while no points are awarded when they have a different position. For example, when a user is neutral and a party disagrees with a statement, no points are awarded. Double points are awarded for issues that have been given extra weight by the user. The best match is the party with most points.

The agreement method used by StemWijzer is straightforward and relatively easy to understand. However, inspired by research in political science in which political parties are
mapped in a so-called political space, many VAAs adopt a different approach: they use the items to create a spatial framework in which voters and parties are positioned (cf. Downs 1957; Enelow and Hinich 1984). The implicit advice of such tests seems to be to cast a vote for the party that is closest to oneself in the political space, analogous with the smallest distance hypothesis in spatial models of voting (see Downs 1957). In the Netherlands this procedure has been adopted by *Kieskompas*, which was developed by André Krouwel, an entrepreneur with an academic background in research on political parties. The company that exploits *Kieskompas* has also developed similar tests for other countries, such as Canada, Sweden and Turkey. Some other VAAs, like the *EU Profiler*, also use an image of a two-dimensional political space in which user and parties or candidates are plotted as one of the ways to present the result.

An example of a VAA that was modelled after the Dutch *Kieskompas*, and for which the method has been documented in some detail (Lobo, Vink and Lisi 2010), is the Portuguese *Bússola Eleitoral* ("electoral compass"). In the run-up to the 2009 national elections this test was completed about 175,000 times, which corresponds with 3 percent of the electorate. The developers identified a socio-economic dimension of left/right as the major basis for party competition in Portugal. Additionally, they formulated a second dimension that resembles the division between Green/Alternative/Libertarian (GAL) and Traditional/Authority/Nationalist (TAN) (cf. Marks, Hooghe, Nelson and Edwards 2006). In total 28 statements were selected, of which 15 linked up with Left/Right and 13 linked up with GAL/TAN. The answer format was a five-point rating scale ranging from “completely agree” to “completely disagree”. The positions of users and parties were determined by taking the average position on each item related to that dimension, using a scale from −2 to +2. Party positions were determined on the basis of election manifestos, debates in parliament, and statements on party websites and in the media. The website also asked users
to rate party leaders and indicate the chance they would ever vote for each party, but answers to these questions were not used for calculating the match.

The match between user and party was determined by Bússola Eleitoral in two ways. First, on the basis of answers to the statements the position of users and parties on both axes of the political space were determined and plotted on the screen (see Figure 1b). The implicit message is that the party closest in this space (Euclidean distance) provides the best match. Second, agreement scores were determined by calculating the average agreement on all 28 statements. Lobo et al. (2010: 168) argue this is actually a better measure: “These scores are more accurate because, hypothetically, it could well be the case that a user occupies the same position on both of the axes, as one of the parties, but that this proximity is based on agreement and disagreement with different statements from the questionnaire.” This is indeed true, although the empirical question remains to what extent in practice both methods lead to different results. One could add that the argument applies in particular to centrist parties. For political parties at the extremes it is logically impossible to occupy the same position without answering statements in a similar way. Taken one step further, one could hypothesize that the method of averaging scores is advantageous for centrist parties (that is, for parties that are centrist in the model used by the VAA).

The largest VAA in Switzerland, smartvote, makes use of a framework that combines eight policy dimensions (see Fivaz and Schwarz 2007; Ladner et al. 2010). This test appeared on internet in 2003, when it was consulted 250,000 times. In 2007 the figure increased to 1.0 million, which corresponds with about forty percent of the Swiss electorate. Smartvote presents the outcome of the test with a bar chart (cf. StemWijzer), a two-dimensional model, and a spider diagram (which they call smartspider). The spider diagram of smartvote comprises eight axes with values ranging between 0 and 100. Each voter and party is positioned on each dimension. The points that represent the positions of a voter are connected
by lines, which results in an area in the diagram. The same is done for parties (or candidates, in the case of *smartvote*). By comparing the areas of voter and party, which can be done in the same diagram, one can see the amount of overlap and hence the degree of match between both (see Figure 1c). If the areas overlap completely, voter and party take identical positions on each dimension and the match is perfect. The less overlap there is between the voter area and the party area, the poorer the match between both. To find out which party provides the best match the graphs for all parties (combined with the voter) have to be compared by the user.

The above overview is not an exhaustive list of all methods employed by VAAs. But we believe that it does capture the essentials of the ways in which VAAs differ with respect to the methods used to determine the match between voter and party and present the advice. Put simply, there are four methods. The first method, which is central in *StemWijzer*, corresponds with counting the proportion of statements where voter and party agree. The best match is provided by the party with whom the voter agrees most often. This method implicitly adopts a multidimensional spatial framework in which each item represents a separate dimension. This method can use different metrics: city block distance, Euclidean distance or the agreement method used in the most recent version of *StemWijzer*. The second method is the one that *StemWijzer* started with, namely by constructing a one-dimensional space, like a left/right continuum, and position voters as well as parties on this scale. The third method, which has been used by *Kieskompas* and its spin-offs, consists of the positioning of voters and parties in a two-dimensional space. Positions of voters and parties are typically determined by averaging the scores on the items that are considered as indicators of that policy dimension. The fourth method also adopts a spatial framework, but distinguishes multiple dimensions. *Smartvote*, which distinguishes eight policy dimensions, is a typical example.
In this paper we compare these spatial models and examine if the use of different spatial models would lead to different advice at the individual level as well as differences across parties at the aggregate level. Additionally, we analyse if the use of a different metric (Euclidean distance, city block distance, agreement method) affects the result.

**Data and method**

*Selection of users and parties in the analysis*

The analysis we present is based on the responses given by the 4.2 million users of the StemWijzer edition for the Dutch parliamentary election in 2010. The thirty items included in this test are listed in Appendix A. The log files of the online application contain information on the positions taken on each of the statements as well as the extra weight allocated by users to each statement. This allowed us to calculate the results under different methods of VAA advice calculation.

Answer profiles that only contained missing answers (‘skip this question’) were excluded. In addition, we excluded about 30,000 advices because we suspect that these have been computer-generated. These represent three cases where one single IP address requested many thousands of advices giving exactly the same answers to the statements each time. Next, we took a random sample of 10,000 cases for further analysis, because analysing the full dataset would be too memory-intensive. While this may introduce some error, a sample of this size will yield results that are almost certainly extremely close to what would have been found using the full dataset. Most users in our sample (90%) answered all 30 statements and very few skipped more than five questions (1%). A majority (77%) made use of the opportunity to select statements that they considered particularly important (mean = 5.1; standard deviation = 4.7).
The 2010 edition of *StemWijzer* by default included the 11 political parties that were represented in the Second Chamber of the Dutch parliament at that time. Voters had the opportunity to deselect any of these parties, while they could also select any of six additional parties for which the application included data. The log files do not contain information on which parties were (de)selected by each user, but the low number of advices for parties that were not included by default (2%) suggests that only few people used this possibility. Because we are interested in the effect of VAA aggregation procedures on the voting advice of users, we opted to include the default selection of parties in our analysis. This presumably most truthfully reflects the advices a majority of users received.

*Alternative methods of calculating voting advice*

We implemented eight methods of calculating a vote advice: a) high-dimensional agreement method (the method that *StemWijzer* used in its 2010 edition), b) high-dimensional city block distance method, c) high-dimensional Euclidean distance method, d) one-dimensional model, e) two-dimensional model, f) three-dimensional model induced from parties’ answers to the statements, g) three-dimensional model incuded from users’ answers to the statements, and h) seven-dimensional ‘spider’ model. The implementation of the first three methods is relatively simple (see Appendix B). One calculates the agreement or distance between the answers of a particular user and those of each party, ignoring missing answers on the part of the user. Extra weights used by the users increase the agreement or distance between users and the party. For the agreement method this results in an agreement score, which is highest for the ‘best match’. The city block and Euclidean distance method result in (weighted) distance scores, which is lowest for the ‘best match’. We included the extra weights voters could put on statements in these models (as this has been standard practices in *StemWijzer* and other
VAAs), but our findings would have been similar if these weights would have not been taken into account.

The one-, two-, and multidimensional models require a method of combining items into issue dimensions. For the one-dimensional model this is a question of determining the direction of each statement: does agreeing imply a left-wing/progressive position or rather a right-wing/conservative position? We determined this on a priori grounds and checked the homogeneity of the resulting issue dimension using Loevinger’s $H$. When looking at the answers of the political parties, this results in a scale with $H = 0.37$, which is low but acceptable. For the users, the homogeneity coefficient is very low ($H = 0.07$), which indicates that in this case for voters a one-dimensional approach is insufficient. Nonetheless, we include this method of calculation in the analysis to demonstrate what its adaptation would mean. In addition, some observers have argued that a one-dimensional left-right model is sufficient in the context of electoral choice (Van der Eijk and Niemöller 1983).

The implementation of the two-dimensional model follows the two-dimensional Kieskompas model as closely as possible. The model consists of a socio-economic left-right dimension and a progressive-conservative (GAL/TAN) dimension (Marks et al. 2006). Each of the thirty statements has been assigned to either of these dimensions, similar to the approach followed by Kieskompas. The resulting scales are not very strong with $H$ coefficients of 0.20 and 0.40 respectively for the parties’ answers, and very weak $H$ coefficients of 0.06 and 0.07 for the user data. Similar results concerning the strength of the Kieskompas model have been found by Otjes and Louwerse (2011: 10-13).

The two-dimensional model has been constructed by selecting the relevant policy dimensions a priori. However, one could argue that first relevant political issues should be selected and that the appropriate spatial model should be induced from the patterns of (party or voter) answers given to these statements (Otjes and Louwerse 2011). We may thus find out
inductively that parties’ answers to the statements can be captured very well by a one-dimensional or two-dimensional spatial model. This type of model was fitted using classical Multidimensional Scaling (MDS). This method uses a Euclidean distance measure between actors (parties or users), based on their answers to the VAA statements, and tries to find a low-dimensional approximation of those distances. We applied this method in two ways, namely once on the basis of party positions and once on the basis of voter positions. The degree to which a low dimensional model accurately represents the distances between parties is measured by Kruskal’s Stress-I statistic. Stress levels below 10 percent are considered acceptable. For the dataset of party responses to the statements a three-dimensional solution was found to be acceptable (Stress = 4.37). The next step was to determine to which dimension each statement was connected, which was determined by regressing parties’ answers to each statement on these three dimensions, a technique called property fitting. We included an item in the dimension that provided the highest beta coefficient for that particular item, provided the R square was larger than .30. In this way 29 out of the 30 items were included in one of the scales (see Appendix B). We use additive scales to construct the model, because this reflects most closely how other VAAs, such as Bússola Eleitoral and Kieskompas, construct their spatial model. An additional advantage of this method is that for users it is more transparent than more sophisticated techniques such as factor analysis or multidimensional scaling. The resulting scales had high $H$ values of 0.71, 0.75 and 0.54 respectively (based on parties’ answers). However, when applied to the users’ answers these scales are not very strong ($H = 0.09, 0.06$ and 0.15).

In an alternative specification users’ answers to the statements were used in an MDS analysis. Thus, instead of a ‘party space’ a ‘voter space’ was constructed. Answer patterns of users proved to be more erratic than those of parties: a three dimensional solution had a stress level of 30%, but including more dimensions only reduced this level very gradually to 13%
for a ten-dimensional model. For reasons of clarity and comparability, we decided to stick to a three dimensional model in this case as well, despite the poor fit. After all, the logic behind a low-dimensional model of party positions is to provide users insight in the different policy stances of parties – presenting a ten-dimensional model would destroy this objective. In addition, Aarts and Thomassen (2008) have also induced three-dimensional spatial models of party positions for the Dutch case (1989-2006), based on voters’ sympathy ratings of parties. The H coefficients for the three dimensions were 0.29, 0.08 and 0.13 respectively, which is (somewhat) better than the $H$ values (for voters) obtained from the party space, as one might expect. Still, the scalability of these items is low. To calculate the resulting vote advice from the multidimensional models, the distance between users and parties were calculated as (unweighted) Euclidean distances.

The last method, a seven-dimensional model that reflects the spider diagram, has been implemented by assigning issues to one of the seven categories that were used in the EU Profiler’s spider diagram (note that smartvote uses eight spider categories, but the principle is the same). The assignment was based on a priori grounds and checked using the homogeneity coefficient $H$. For the answers provided by the parties, the $H$ values for each of the Smart Spider issues were over 0.3, except for welfare state politics. Although the coefficient for the latter category could be improved by changing the direction of some of the items, this would run contrary to the substantive meaning of the category. The a priori approach seems to fit most closely to how smartvote and EU Profiler construct their spider models. Users’ and parties’ answers to the statements were recoded and summed up, so that a score of 100 on a particular axis indicates complete agreement and 0 complete disagreement. The total distance between voters and parties was calculated as the city block distance, which seems to fit most closely to the way a spider diagram represents the vote advice.
Three measures to compare the models

We use three different measures to compare the advices stemming from the alternative spatial models. The first measure focuses on the party that provides the best match and indicates how often two given methods provided the same ‘best match’. If someone received the advice “Freedom Party” using the city block method as well as the Euclidean method, this constitutes a full match. There are also cases where the ‘best match’ was a tie between two or more parties. When at least one party was among the best matches in both methods this was regarded as a partial match. All other cases were treated as ‘no match’.

The second measure focuses on the overall degree of match between the user and each individual party. To capture the similarity of the overall advice, we calculated a correlation coefficient between the match scores of two methods for each individual party. For example, if there is a perfect linear relationship between the Labour Party scores according to the agreement method and the city block method, the correlation coefficient equals one. To estimate the overall similarity between two methods we take the means of these correlations across the 11 parties.

The third measure looks at the aggregate number of advices for parties. Some methods may divide advices more evenly over parties, while other methods may favour specific parties. Furthermore, it is possible that particular parties may “benefit” from a particular method. Therefore, we look at the aggregate number of advices for each party according to each method (where tied advices were divided equally over the parties concerned).

TABLE 1 ABOUT HERE
Results

Effects at the individual level

The eye-catching element of any VAA result page is the ‘best match’ between the user and a party. Table 1 displays the similarity between the ‘best matches’ that users would have received under different methods of calculating the VAA result. The similarity between the agreement, city block and Euclidean methods is high: for each dyad, over 80 percent of users would receive exactly the same ‘best match’. Given the fact that each of these methods treats the statements as independent and adds up differences between parties and voters on each statement, this is in line with our expectation. Nonetheless, the advices are not exactly the same. Because the ‘best match’ of one method is in many cases only slightly better than the ‘second best match’, small differences in the calculation method may affect which party appears on top of the list.

The agreement method does not always provide the user with a single ‘best match’. In about 20% of the cases the ‘best match’ was a tie between two or more parties. The percentage of users with a tied advice was lower for the city block and Euclidean methods, 12% and 9% respectively, and almost non-existent for any of the spatial methods (1% to 3%). The ties influenced the degree of similarity between advice based on agreement scores, city block metric and Euclidean distance: in most cases where no full match was observed, a partial match existed (for 99 percent of the users a partial match between agreement and city block method was observed, while for agreement method and Euclidean distance the corresponding figure was 94 percent). Because ties were uncommon in the one-dimensional and multi-dimensional spatial models, the comparisons between those are not strongly affected by the potential presence of ties.
The similarity between the various spatial methods and the high-dimensional approaches is quite low. This is certainly true for the one-dimensional model: only 6% of users would receive an identical advice on the basis of the one-dimensional model and high-dimensional agreement method (and for only 5% there is a partial match). This result alone already shows that it greatly matters which underlying model is used. The other spatial approaches show a somewhat higher similarity with the agreement method, but also reveal strong differences between methods. The two-dimensional model provides the same best match as the agreement method in 23% of the cases (with 9% partially matching). The three-dimensional space that was constructed on the basis of parties’ positions gives the same advice as the agreement method in only 13% of the cases (while for 9% it gives a partial match). The three-dimensional voter model does the same for 35% of the cases. The seven-dimensional ‘spider’ model is somewhat more congruent with the agreement method: 41% of users receive a similar best match, while another 13% had a partial match between both methods. Thus, as the dimensionality of the underlying model goes down, the advices get more different from the agreement method.

Apart from the difference between high-dimensional and low-dimensional models, there is also a difference within the group of low-dimensional (spatial) models. For example, the ‘best matches’ provided by the three-dimensional model based on parties’ positions is only in 19% of the cases the same as the advice provided by the three-dimensional model based on voters’ positions. It matters how a particular model is constructed.

While the ‘best match’ is informative of the different results of the various methods, it only looks at one particular aspect: the top of the stack. It is possible that these shifts in
advice stem from rather small differences in the scores of parties. Therefore, we have calculated how similar advices are by looking at the correlation between party scores for any two methods (see Table 2). The figures follow the pattern that was observed in Table 1. The agreement, city block and Euclidean method show largely similar advices, while the average correlation between one-dimensional and other methods of calculating the voting advice is moderate. For example, the correlation between the scores based on the agreement method and the three-dimensional spatial model amounts to about 0.6. This means that the amount of explained variance amounts to less than 40 percent. So the main observation is that the correlations are rather low, given the fact that the different methods all aim at the same outcome: determining how well the policy positions of a party matches with opinions of a voter. These figures underline the conclusion based on the previous measure: the method strongly affects to what extent users and parties are perceived as holding similar policy positions.

**TABLE 3 ABOUT HERE**

*Effects at the aggregate level*

The final question is whether the differences in advice at the individual level translate into differences at the aggregate level. Does the number of advices that a party receives, so to speak, differ between methods, or do the differences at the individual level cancel each other out? Table 3 shows how the different methods translate into total numbers of advices for each party, while also presenting the actual election outcome.

The agreement method that was used by *StemWijzer* produced 34% advices for the Freedom Party (PVV) led by Geert Wilders, which therefore was the party that most often matched best with users’ policy preferences. The figure for the Labour Party (PvdA) was
19%, while all other parties score below 10%. In a sense the high figure for Wilders’ party is not surprising, because research has shown that a large group of voters combines left-wing policy preferences on the socio-economic left-right dimension with right-wing policy preferences on the cultural left-right dimension (Van der Brug and Van Spanje 2009). The Freedom Party approximates this position most closely. The aggregate outcome for the 11 parties would not have been very different would StemWijzer have used city block metric or Euclidean distances: the percentages for each party are nearly identical. One exception is the orthodox Protestant party SGP, which received 2.8% of best matches in our sample according to the agreement method, but would have more than doubled that if the Euclidean distance method would have been adopted. Yet, on the whole the differences are limited.

FIGURE 2 ABOUT HERE

If we compare the high-dimensional agreement method with the other spatial models, sizeable differences can be observed. The one-dimensional model provides the most deviating outcome. This method would have led to a rather different set of advices: over 53% of best matches would concern the SGP and over 27% the other small Protestant party Christian Union. This remarkable outcome can be explained by looking at the distribution of parties and voters on this single dimension, which is displayed in Figure 2. The top half of the figure displays the distribution of the users in our sample by means of a violin plot, a combination of a box-plot displaying the mean, quartiles and (truncated) range of values, and a density plot. It shows that 50% of the users scored between -.167 and +.167 on the continuum, which ranges from -1 to +1. The parties’ positions are located much more towards the extremes and in two clusters: one left-wing and one right-wing. Thus, a large majority of users has a left-right position somewhere in between the left block and the right
block. As a result, the parties in either block that are located closest to the centre, receive
these advices. For the left-wing block this is the Christian Union and for the right-wing block
the SGP. The paradoxical result is that the parties which fit the left-right framework most
poorly, the two orthodox Protestant parties (Pellikaan et al. 2003), would benefit most strongly from adapting this method for calculating the VAA result.

The multi-dimensional models suffer largely from the same problem as the one-
dimensional model: voters are clustered in the centre and therefore parties which are close to
the centre do very well in terms of ‘best matches’. Which parties are close to the centre,
depends strongly on the model adapted. In the two-dimensional model, Freedom Party
(PVV), Christian Union (CU) and Democrats 66 (D66) are close to the centre, for the Party
MDS model it is the Socialist Party (SP), Christian Union (CU) and Proud of the Netherlands
(TON), while for the Voter MDS model, the Freedom Party (PVV) and Labour (PvdA) are
located more towards the centre. This is reflected in the number of best matches for these
parties. The seven-dimensional ‘spider’ representation also suffers from this problem, but to a
more limited extent. Compared to the agreement method, one party scores much better in this
model, Trots op Nederland (TON), while most of the other parties do slightly worse.

It is clear that the election result does not accurately reflect the advices of StemWijzer,
or vice versa. This is well illustrated by the fact that the election winner, Liberal Party
(VVD), received only 7% of the advices. This suggests that many voters did not vote for the
party that seemed closest in terms of policy, or at least not in terms of the 30 policy items
included in this test. The figures indicate that with another metric or another spatial model the
situation would have been the same. None of the methods would have resulted in an
aggregate level outcome that matches the actual election result. This is not at all worrisome,
however, because the aim of VAAs is not to predict election outcomes and voters may have

21
many other reasons to support a particular party (e.g., social influence, habit, religious background, government approval, party leader effects, and so on).

Conclusions

Voting Advice Applications (VAAs) have become an important tool for voters to find out how well their policy preferences match with the positions taken by political parties. Given their widespread use, the potential impact on election outcomes is substantial. Against this background we have examined the relevance of the spatial framework that VAAs adopt on the basis of the answers by users of StemWijzer, a Dutch VAA that was consulted by about one third of the voters in the run-up to the 2010 national elections. In our analysis we compared the method used by the original application, which corresponds with a simple count of the proportion of statements where user and party took the same position, with spatial models that are central in other VAAs. More specifically, we compared the StemWijzer method with a one-dimensional, two-dimensional, three-dimensional and seven-dimensional spatial model. Additionally, we analysed the impact of the type of metric used to calculate the match: Euclidean distance, City block distance, and StemWijzer’s agreement method.

The analyses have clearly shown that the design matters. Up to 90 percent of users would have received a different advice if another method for calculating the match between voter and party had been adopted. Although in some cases this may have been due to the fact that the match with two or three parties was fairly similar, additional analyses showed that the main reason for these discrepancies is that the correlation between measures of the match across methods is only moderately strong. So the degree of match between voter and party as calculated by a VAA not only depends on the positions both take on the policy items included in the test, but also on the method used to calculate the match. Also in terms of the aggregate
number of ‘best matches’ for each party, there are large differences. For example, whereas only 3 per cent of the users were told that their policy preferences matched best with the Dutch Reformed Party (SGP), a small orthodox Protestant party, in a one-dimensional model more than half of the users would have been perceived as closest to this party. With two-dimensional and multi-dimensional models the differences were not as extreme as in this case, but sizeable differences in the number of advices for each party were still observed.

How can these findings be taken into account when VAA makers choose the method for calculating the vote advice? There are two elements of the design that merit attention. The first is the way in which the advice is presented, especially whether a particular party is singled out as providing the best match. One argument to present a result screen that indicates which single party provides the best match is that this may be what the user desires and what makes VAAs appealing. After all, in the election only one party or candidate can be given the vote and the decision making is therefore not much facilitated by details about a rank ordering or the precise differences between the best match and the second-best match, is sometimes argued. However, the downside of a clear advice for a single party is that this may suggest more precision and objectivity than what can be substantiated. Previous research by Walgrave et al. (2009) demonstrated that statement selection by VAAs has a substantial impact on the nature of the advice. We have shown that with the same set of statements, the method to calculate the match between voter and party also has a strong influence on the outcome of the test. As long as there is no agreement about the proper selection of statements and the best method to integrate these into a voting advice, an advice in the format of the name of a single party would not do justice to the small differences between parties and the strong sensitivity of the advice in relation to the method employed. So presenting the full rank ordering of parties with the degree of agreement reflected in a bar chart or a two-
dimensional spatial model in which user and parties are represented, is a more appropriate way to present the advice than merely providing the name of a single party.

The second implication of our findings is that the use of a low-dimensional model to capture and compare parties’ and users’ policy positions is not unproblematic. The spatial models used in this paper do not fit very well with voters’ answers to statements. A similar finding has been reported by Kleinnijenhuis and Krouwel (2009) who studied the two-dimensional model underlying EU Profiler and found that eight (varimax rotated) principal components were necessary to explain 58% of the variance in voters’ answers. This suggests that the problem of constructing a low-dimensional model is not merely a result of the relatively complicated nature of the Dutch party system, although the presence of a large number of (relevant) parties will certainly not help to make these models more concise. The question remains, however, whether the bad fit in the spatial dimensional models has to do with the statement selection and use of the answer scale. The statements included in StemWijzer have not been selected with a spatial model in mind. Moreover, the fact that statements were selected in such a way that any pair of parties can be adequately distinguished might have influenced the fact that the items cannot be reduced to a limited number of dimensions. Furthermore, by encouraging or even forcing parties to take position either in favour or against an issue statement, parties are clustered more towards the end of policy dimensions, while voters are concentrated in the centre (users chose the ‘neither’ option in 11.2% of cases, while the parties only selected it in 2.4% of cases). So if a spatial model is to be used, it may be preferable to use a more fine-grained mechanism for capturing answers to statements, such as the four- or five-point scales that smartvote, EU Profiler, Kieskompas and Bússola Eleitoral use. In addition, VAA designers may opt to use various alternative methods to calculate and present the vote advice side-by-side, examples of which
are *smartvote* and *EU Profiler*, which used both a match list, two-dimensional model and a spider model (Trechsel and Mair 2009: 9-11).

More important, however, is that VAA designers should check whether the spatial models they employ are indeed valid depictions of the party competition and whether they make sense as a tool to position voters (Otjes and Louwerse 2011). The effect of different models can be taken into account in the VAA design process. Walgrave et al. (2009) suggest that VAA designers should select statements based on a mass survey which puts potential statements to the test. The same applies to the design and use of particular spatial models. This way, VAA designers can decide *ex ante* which statement fits which model best and how this will probably affect the result. While we argue there is no golden standard for deciding what is the ‘best’ selection of items or method for calculating the result, this will at least allow designers to estimate the stability of the advice across methods.
References


# Appendix A: Statements in *StemWijzer 2010*

<table>
<thead>
<tr>
<th>Statement</th>
<th>% agree</th>
<th>% disagree</th>
<th>% extra weight</th>
<th>2-D party</th>
<th>3-D party</th>
<th>3-D voter</th>
<th>7-D</th>
</tr>
</thead>
<tbody>
<tr>
<td>The government has to cut spending with billions. In 2015 at latest the national budget deficit should have been resolved.</td>
<td>41</td>
<td>36</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>People who earn very much should pay more taxes than they do now.*</td>
<td>59</td>
<td>33</td>
<td>27</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>The amount of child benefits should become dependent on someone’s earnings.*</td>
<td>70</td>
<td>26</td>
<td>24</td>
<td>1</td>
<td>1</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>Shopkeepers should be able to determine themselves if they open their shop on Sunday.*</td>
<td>76</td>
<td>20</td>
<td>10</td>
<td>2</td>
<td>-</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>People who become unemployed receive unemployment benefits. The amount should be increased in the first months, but the total period of benefits should be strongly reduced.</td>
<td>44</td>
<td>39</td>
<td>19</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>From the age of 65 onwards everybody receives elderly benefits. This should remain like it is.*</td>
<td>71</td>
<td>22</td>
<td>33</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>6</td>
</tr>
<tr>
<td>It should become easier for employers to fire employees.</td>
<td>30</td>
<td>58</td>
<td>20</td>
<td>1</td>
<td>3</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>All Islamic schools in the Netherlands should be closed.</td>
<td>34</td>
<td>51</td>
<td>13</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Extra money is needed for education and therefore the government should abolish student funding. From now on students should borrow money for their study.</td>
<td>19</td>
<td>72</td>
<td>26</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td>6</td>
</tr>
<tr>
<td>The government pays for public television channels “Nederland 1, 2 and 3.” At least on of the channels should disappear.</td>
<td>50</td>
<td>38</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>The government should strongly reduce subsidies for art and culture.</td>
<td>55</td>
<td>32</td>
<td>11</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>In the Netherlands no new mosques may be built.</td>
<td>31</td>
<td>53</td>
<td>14</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>Someone who owns a house receives tax benefits. This arrangement should remain like it is.</td>
<td>67</td>
<td>23</td>
<td>39</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Too many people with high income live in social housing. Therefore they should pay more rent for the same house than people with lower income.*</td>
<td>57</td>
<td>36</td>
<td>15</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>The government should spend extra money on new and wider motorways.</td>
<td>46</td>
<td>38</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Possession of cars should become cheaper and use of cars more expensive. Therefore a mileage tax should be introduced.*</td>
<td>43</td>
<td>45</td>
<td>20</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>Statement</td>
<td>56</td>
<td>37</td>
<td>18</td>
<td>2</td>
<td>1</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>--------------------------------------------------------------------------</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>-----</td>
<td>----</td>
<td>-----</td>
</tr>
<tr>
<td>Everybody should be an organ donor, unless written objection has been made.*</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Assistance with suicide should be possible under specific conditions.*</td>
<td>83</td>
<td>11</td>
<td>18</td>
<td>2</td>
<td>2</td>
<td>-</td>
<td>7</td>
</tr>
<tr>
<td>There should be more competition between care institutions (free market system).</td>
<td>35</td>
<td>51</td>
<td>12</td>
<td>1</td>
<td>3</td>
<td>1 &amp;</td>
<td>6</td>
</tr>
<tr>
<td>The queen has several political tasks. This should be changed, so she has only ceremonial tasks.*</td>
<td>36</td>
<td>49</td>
<td>8</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>To give voters more influence they should have the possibility to reverse decisions from national politics by referendum.*</td>
<td>59</td>
<td>28</td>
<td>12</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>-</td>
</tr>
<tr>
<td>People who are employed by the government should not be allowed to wear head scarves at work.</td>
<td>47</td>
<td>43</td>
<td>14</td>
<td>2</td>
<td>2</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>There should be much more severe punishment for people who use violence.</td>
<td>88</td>
<td>6</td>
<td>35</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>The government should make the growth and sales of soft drugs legal.*</td>
<td>48</td>
<td>39</td>
<td>11</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>Organizers of sports events should pay themselves for the use of police force.*</td>
<td>57</td>
<td>30</td>
<td>13</td>
<td>1</td>
<td>2</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>The Netherlands should spend much less on developmental aid.</td>
<td>54</td>
<td>33</td>
<td>19</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>The government should strongly cut spending on defense.*</td>
<td>53</td>
<td>28</td>
<td>14</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>-</td>
</tr>
<tr>
<td>There should be a prohibition on very large stables for livestock (mega stables).*</td>
<td>54</td>
<td>26</td>
<td>10</td>
<td>2</td>
<td>1</td>
<td>2</td>
<td>5</td>
</tr>
<tr>
<td>To improve the environment, taxes on meat should be strongly increased.*</td>
<td>22</td>
<td>65</td>
<td>8</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>The Netherlands should be allowed to build a second nuclear plant.</td>
<td>42</td>
<td>43</td>
<td>13</td>
<td>2</td>
<td>1</td>
<td>3</td>
<td>5</td>
</tr>
</tbody>
</table>

Note: Statements marked with * are considered left-wing positions in the one-dimensional model, the other statements are considered right-wing. Columns three to seven indicate to which dimension each statement belongs.
Appendix B: Calculation of voting advice according to the different methods

Agreement method

\[
A_{rs} = \sum_{i=1}^{30} w_{ri} a_{rsi}
\]

Where \(a_{rsi}\) equals one when user \(r\) and party \(s\) provided the same answer to statement \(i\), zero otherwise, and \(w\) equals two when the user put extra weight on an issue and 1 otherwise, with \(i = 1\) to 30 statements.

Note: The Agreement method may also be considered a sixty (30 times 2) dimensional model, in which each statement is translated into two dimensions, e.g. agree = (½,0), disagree = (-½,0) and neutral = (0,½). Note that on a single statement, the city block distance between ‘agree’ and ‘neutral’ is equal to one, just as the distance between ‘disagree’ and ‘agree’. The total (dis)agreement score is the city block distance between the parties’ and voters’ positions on the sixty dimensions.

City block distance method

\[
CB_{rs} = \sum_{i=1}^{30} w_{ri} |p_{ri} - p_{si}|
\]

Where \(p_{ri}\) is the voters’ position on issue \(i\) and \(p_{si}\) the party’s position. The weight \(w_{ri}\) equals two when the user put extra weight on an issue and 1 otherwise, with \(i = 1\) to 30 statements.

Euclidean distance method

\[
E_{rs} = \sqrt{\sum_{i=1}^{30} w_{ri} (p_{ri} - p_{si})^2}
\]

Where \(p_{ri}\) is the voters’ position on issue \(i\) and \(p_{si}\) the party’s position. The weight \(w_{ri}\) equals two when the user put extra weight on an issue and 1 otherwise, with \(i = 1\) to 30 statements.
Figure 1: Result screens of Voting Advice Applications (VAAs)

a. Dutch VAA *StemWijzer*

![Dutch VAA StemWijzer](image)

b. Portuguese VAA *Bússola Eleitoral*

![Portuguese VAA Bússola Eleitoral](image)

c. Swiss VAA *smartvote*

![Swiss VAA smartvote](image)
Figure 2: Parties’ and users positions in the one-dimensional model

Note: Voters’ positions are based on sample of 10,000 VAA users. The plot is a violin plot, which combines a boxplot and a density plot.
Table 1: Percentage of users who received the same ‘best match’ by two methods

<table>
<thead>
<tr>
<th></th>
<th>Agreement</th>
<th>City block</th>
<th>Euclidean</th>
<th>One-dimensional</th>
<th>Two-dimensional</th>
<th>Three-dimensional (Party MDS)</th>
<th>Three-dimensional (Voters MDS)</th>
<th>Seven-dimensional (‘Spider’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td>100.0</td>
<td>85.2</td>
<td>83.0</td>
<td>6.4</td>
<td>23.2</td>
<td>13.0</td>
<td>35.2</td>
<td>41.3</td>
</tr>
<tr>
<td>City block</td>
<td>85.2</td>
<td>100.0</td>
<td>94.7</td>
<td>7.7</td>
<td>24.6</td>
<td>16.0</td>
<td>37.2</td>
<td>43.7</td>
</tr>
<tr>
<td>Euclidean</td>
<td>83.0</td>
<td>94.7</td>
<td>100.0</td>
<td>9.1</td>
<td>25.0</td>
<td>16.7</td>
<td>37.6</td>
<td>43.9</td>
</tr>
<tr>
<td>One-dimensional</td>
<td>6.4</td>
<td>7.7</td>
<td>9.1</td>
<td>100.0</td>
<td>19.9</td>
<td>18.4</td>
<td>12.4</td>
<td>9.5</td>
</tr>
<tr>
<td>Two-dimensional</td>
<td>23.2</td>
<td>24.6</td>
<td>25.0</td>
<td>19.9</td>
<td>100.0</td>
<td>22.2</td>
<td>29.8</td>
<td>26.9</td>
</tr>
<tr>
<td>Three-dimensional (Party MDS)</td>
<td>13.0</td>
<td>16.0</td>
<td>16.7</td>
<td>18.4</td>
<td>22.2</td>
<td>100.0</td>
<td>17.2</td>
<td>22.1</td>
</tr>
<tr>
<td>Three-dimensional (Voters MDS)</td>
<td>35.2</td>
<td>37.2</td>
<td>37.6</td>
<td>12.4</td>
<td>29.8</td>
<td>17.2</td>
<td>100.0</td>
<td>38.4</td>
</tr>
<tr>
<td>Seven-dimensional (‘Spider’)</td>
<td>41.3</td>
<td>43.7</td>
<td>43.9</td>
<td>9.5</td>
<td>26.9</td>
<td>22.1</td>
<td>38.4</td>
<td>100.0</td>
</tr>
</tbody>
</table>
Table 2: Mean correlation between the advices of two methods (Pearson’s R)

<table>
<thead>
<tr>
<th></th>
<th>Agreement</th>
<th>City block</th>
<th>Euclidean</th>
<th>One-dimensional</th>
<th>Two-dimensional</th>
<th>Three-dimensional (Party MDS)</th>
<th>Three-dimensional (Voters MDS)</th>
<th>Seven-dimensional (‘Spider’)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Agreement</td>
<td>1.00</td>
<td>0.91</td>
<td>0.79</td>
<td>0.51</td>
<td>0.51</td>
<td>0.58</td>
<td>0.57</td>
<td>0.60</td>
</tr>
<tr>
<td>City block</td>
<td>0.91</td>
<td>1.00</td>
<td>0.97</td>
<td>0.58</td>
<td>0.58</td>
<td>0.66</td>
<td>0.66</td>
<td>0.69</td>
</tr>
<tr>
<td>Euclidean</td>
<td>0.79</td>
<td>0.97</td>
<td>1.00</td>
<td>0.56</td>
<td>0.56</td>
<td>0.64</td>
<td>0.66</td>
<td>0.69</td>
</tr>
<tr>
<td>One-dimensional</td>
<td>0.51</td>
<td>0.58</td>
<td>0.56</td>
<td>1.00</td>
<td>0.80</td>
<td>0.87</td>
<td>0.68</td>
<td>0.60</td>
</tr>
<tr>
<td>Two-dimensional</td>
<td>0.51</td>
<td>0.58</td>
<td>0.56</td>
<td>0.80</td>
<td>1.00</td>
<td>0.75</td>
<td>0.62</td>
<td>0.62</td>
</tr>
<tr>
<td>Three-dimensional (Party MDS)</td>
<td>0.58</td>
<td>0.66</td>
<td>0.64</td>
<td>0.87</td>
<td>0.75</td>
<td>1.00</td>
<td>0.67</td>
<td>0.64</td>
</tr>
<tr>
<td>Three-dimensional (Voters MDS)</td>
<td>0.57</td>
<td>0.66</td>
<td>0.66</td>
<td>0.68</td>
<td>0.62</td>
<td>0.67</td>
<td>1.00</td>
<td>0.67</td>
</tr>
<tr>
<td>Seven-dimensional (‘Spider’)</td>
<td>0.60</td>
<td>0.69</td>
<td>0.69</td>
<td>0.60</td>
<td>0.62</td>
<td>0.64</td>
<td>0.67</td>
<td>1.00</td>
</tr>
</tbody>
</table>

Note: These are the means of correlations for each party.
### Table 3: Percentage of best matches for each party, by method

<table>
<thead>
<tr>
<th>Party</th>
<th>Agreement</th>
<th>City block</th>
<th>Euclidean</th>
<th>One-dimensional</th>
<th>Two-dimensional</th>
<th>Three-dimensional (Party MDS)</th>
<th>Three-dimensional (Voters MDS)</th>
<th>Seven-dimensional (‘Spider’)</th>
<th>Election result 2010</th>
</tr>
</thead>
<tbody>
<tr>
<td>Liberal Party (VVD)</td>
<td>6.8</td>
<td>7.0</td>
<td>6.7</td>
<td>1.1</td>
<td>0.7</td>
<td>0.6</td>
<td>2.9</td>
<td>6.8</td>
<td>20.5</td>
</tr>
<tr>
<td>Labour Party (PvdA)</td>
<td>18.8</td>
<td>17.7</td>
<td>17.0</td>
<td>1.2</td>
<td>2.9</td>
<td>0.8</td>
<td>13.2</td>
<td>18.8</td>
<td>19.6</td>
</tr>
<tr>
<td>Freedom Party (PVV)</td>
<td>34.0</td>
<td>32.8</td>
<td>31.8</td>
<td>3.5</td>
<td>36.6</td>
<td>0.6</td>
<td>43.3</td>
<td>34.0</td>
<td>15.5</td>
</tr>
<tr>
<td>Christian Democrats (CDA)</td>
<td>4.8</td>
<td>3.9</td>
<td>3.4</td>
<td>8.8</td>
<td>1.1</td>
<td>0.0</td>
<td>1.2</td>
<td>0.4</td>
<td>13.6</td>
</tr>
<tr>
<td>Socialist Party (SP)</td>
<td>9.8</td>
<td>9.7</td>
<td>9.1</td>
<td>1.2</td>
<td>6.3</td>
<td>20.3</td>
<td>8.4</td>
<td>9.8</td>
<td>9.8</td>
</tr>
<tr>
<td>Democrats 66 (D66)</td>
<td>3.5</td>
<td>4.5</td>
<td>5.3</td>
<td>1.2</td>
<td>12.7</td>
<td>1.5</td>
<td>3.9</td>
<td>2.2</td>
<td>7.0</td>
</tr>
<tr>
<td>Green Left</td>
<td>3.6</td>
<td>3.1</td>
<td>2.9</td>
<td>0.5</td>
<td>0.4</td>
<td>0.5</td>
<td>0.7</td>
<td>1.9</td>
<td>6.7</td>
</tr>
<tr>
<td>Christian Union</td>
<td>3.4</td>
<td>3.0</td>
<td>2.7</td>
<td>27.2</td>
<td>27.8</td>
<td>27.0</td>
<td>5.1</td>
<td>3.7</td>
<td>3.2</td>
</tr>
<tr>
<td>Dutch Reformed Party (SGP)</td>
<td>2.7</td>
<td>4.1</td>
<td>6.0</td>
<td>52.7</td>
<td>5.0</td>
<td>3.9</td>
<td>9.2</td>
<td>6.1</td>
<td>1.7</td>
</tr>
<tr>
<td>Party for the Animals</td>
<td>5.1</td>
<td>6.2</td>
<td>6.7</td>
<td>0.3</td>
<td>0.5</td>
<td>3.4</td>
<td>3.5</td>
<td>4.4</td>
<td>1.3</td>
</tr>
<tr>
<td>Proud of the Netherlands (TON)</td>
<td>7.9</td>
<td>8.8</td>
<td>9.0</td>
<td>2.7</td>
<td>6.4</td>
<td>42.6</td>
<td>9.1</td>
<td>26.8</td>
<td>0.6</td>
</tr>
</tbody>
</table>

Note: in cases where two or more parties were tied in first place, the advice was divided evenly over the parties concerned.
Notes

1 In earlier days its name was *Stichting Burgerschapskunde* and recently the institute has become part of *ProDemos: Huis voor Democratie en Rechtsstaat*.

2 The information about *StemWijzer* in this section is largely based on a chapter by De Graaf (2010). The same method has been applied in Germany, as discussed by Marschall and Schmidt (2010). The user statistics reported, as well as the log files of the 2010 edition, have been provided to us by the developer of *StemWijzer*, Instituut voor Publiek and Politiek in Amsterdam. We are grateful to Jochum de Graaf for arranging this.

3 These figures concern the number of times that the online test was taken. Presumably, there were individuals who completed the test more than once, so the number of voters who took the test may be somewhat lower than these figures suggest. However, data from the *Dutch Parliamentary Election Studies* 2006 and 2010 indicate that in both years about forty percent of those who cast their ballot had consulted a VAA.

4 Some earlier versions of StemWijzer made use of Euclidean distance, but at some point this was changed into city block distance. Source: personal conversation with Jochum de Graaf from Instituut voor Publiek en Politiek, Amsterdam, 25 May 2011. See also Marschall and Schmidt (2010).

5 The precise method used to calculate the agreement scores is not described by Lobo et al. (2010). We presume that distances between voter and party on each individual item were calculated and then added for all 28 items, leading to a scale ranging between 0 (full agreement on all items) and 112 (maximum disagreement on all items: 28*4).